

GENERACIÓN DE ESTRUCTURAS COMPLEJAS A PARTIR DE MARCADORES DISCURSIVOS EN ESPAÑOL

PRADA NICOLETTI, Juan José

Instituto de Computación, Facultad de Ingeniería
Universidad de la República – Uruguay
prada@fing.edu.uy

Fecha de recepción:

6 de marzo de 2012

Fecha de revisión:

31 de mayo de 2012

Fecha de aceptación:

20 de junio de 2012

Resumen: Este trabajo muestra un análisis del discurso basado en encontrar determinadas unidades textuales denominadas marcadores discursivos en textos irrestrictos en español. Los mismos se comportan como operadores que añaden estructura al texto, de manera que los argumentos de estos operadores son los segmentos de texto que entran en la relación determinada por el marcador. Se propone una representación del texto en una estructura de árbol identificando distintos tipos de nodos. En este artículo se presenta un análisis general del problema, el reconocimiento y la representación de las estructuras mediante el formalismo de reglas contextuales y XML. Así mismo, se exponen algunos resultados producto de una implementación para el marcador *además*.

Palabras clave: Marcadores discursivos – procesamiento del lenguaje natural

Abstract: This paper shows an analysis of discourse based on finding certain textual units called discourse markers in unrestricted texts in Spanish. They behave as operators that add structure to text so that the arguments of these operators are entering text segments in the ratio determined by the marker. A text representation of a tree structure identifying different types of nodes is proposed. This article presents an analysis of the problem, the recognition and representation of structures using the formalism of contextual rules and XML. It also shows some results of an implementation product for the marker as well.

Keywords: Discourse markers – natural language processing

Los marcadores del discurso en español y en chino mandarín, Francisco J. RODRÍGUEZ MUÑOZ (ed.), Almería: Universidad de Almería, 2012 (= *Philologica Urcitana. Revista Semestral de Iniciación a la Investigación en Filología*, vol. 7, Septiembre 2012, ISSN: 1989-6778), 45-57

1. INTRODUCCIÓN

En este artículo mostramos cómo es posible representar un texto en un formato manipulable –XML– cuyo centro son los marcadores discursivos. Está basado en el trabajo de Prada (2001) sobre la detección y el marcado de ciertos elementos presentes en el discurso escrito a los que denominados *marcadores discursivos*. Con este término nos referimos a unidades lingüísticas que establecen relaciones entre segmentos textuales. Son, por ejemplo: *en primer lugar, resumiendo, sin embargo, además*, etc. Su cometido es fundamentalmente el de guiar y ordenar los procesos de interpretación asociados a la comprensión de un texto. Se pueden interpretar como *operadores* que añaden estructura al discurso y donde sus argumentos son aquellos segmentos de texto que entran en la relación determinada por el marcador.

Es interesante el estudio de estos signos lingüísticos, ya que permiten establecer relaciones entre las distintas oraciones y párrafos a lo largo de un discurso y generar estructuras más complejas que vinculan varias oraciones.

En la segunda sección, presentamos las características de los marcadores y algunos de los problemas que surgen al estudiarlos. También prestamos atención a otras propuestas de análisis del discurso en cuanto a la búsqueda de relaciones con puntos de contacto con esta propuesta. En la tercera sección, nos ocupamos de la representación basada en la ubicación de los marcadores a lo largo del texto y las reglas para su determinación. En la cuarta sección, planteamos un caso específico de estudio: el marcador *además*. En la quinta sección, mostramos algunos resultados logrados en la implementación. Finalmente, en la sexta sección, enumeramos las conclusiones principales, así como otros trabajos y aplicaciones vinculadas con la temática.

2. LOS MARCADORES DEL DISCURSO

2.1. Características generales

La siguiente es una definición de marcador del discurso extraída de Martín Zorraquino y Portolés (1999: 4057):

Los marcadores del discurso son unidades lingüísticas invariables, no ejercen una función sintáctica en el arco de la predicación oracional –son pues, elementos marginales– y poseen un cometido coincidente en el discurso: el de guiar, de acuerdo con sus propiedades morfosintácticas, semánticas y pragmáticas, las inferencias que se realizan en la comunicación.

El hecho de guiar las inferencias en el proceso comunicativo está determinando, de algún modo, que se establezca un *conjunto de relaciones* entre distintas partes del texto, construyendo estructuras más complejas vinculadas, precisamente, por estas unidades. Las

relaciones pueden ser de distinta índole: justificación de razonamientos, inferencias, contrariedad de enunciados, secuencias ordenadas de frases, etc. Este trabajo se apoya en la clasificación propuesta en Martín Zorraquino y Portolés (1999). Allí se distinguen cinco grupos que incluyen: *estructuradores*, *conectores*, *reformuladores*, *operadores argumentativos* y *conversacionales*. Por lo general, se trata de relaciones binarias y entre segmentos textuales adyacentes, aunque también existen unarias e incluso compuestas. En el Ejemplo (1), *en primer lugar* es un marcador que actúa sobre un miembro (la oración que lo contiene). Sin embargo, este término se suele emplear cuando se pretende mostrar un orden en una secuencia o razonamiento, actuando en forma conjunta con otros marcadores tales como *en segundo lugar*, *en tercer lugar*, *finalmente*; conformando así un operador compuesto.

- (1) Respecto de variables que afectan directamente la conducta tecnológica de las empresas industriales. **En primer lugar**, una devaluación aumentaría los costos de importación vigentes, tanto de bienes de capital como de insumos, e incrementaría la carga de las deudas pendientes en divisas. **En segundo lugar**, la devaluación también reduciría los ingresos locales respecto a los de los países del extranjero. **En tercer lugar**, las restricciones a la importación.

Existen otros signos lingüísticos como la puntuación que pueden incidir tanto en el significado del marcador como en su alcance, e incluso en la presencia o no del mismo. Esto se puede observar en el Ejemplo (2).

- (2) a. Por tanto estudiar le permitió ocupar un alto cargo en la empresa.
b. Por tanto, estudiar le permitió ocupar un alto cargo en la empresa.

En (2a) el término *Por tanto* no puede considerarse conector, sino complemento circunstancial tematizado: *por tanto estudiar*; incluso lo usual sería que estuviera colocado al final de la oración. En (2b), la coma permite definir a *por tanto* como elemento de conexión.

2.2. Puntos de contacto

Como ya hemos expresado, se trata de encontrar relaciones estructurales basándonos en el reconocimiento de determinadas unidades. Este punto de vista es diferente al planteado en la *Rhetorical Structure Theory* de Mann y Thompson (1988) donde no siempre hay unidades textuales visibles específicas que determinen la presencia de una relación, sino que esta se produce en función de la interpretación que el analista hace del texto. Este hecho implica que sea un tanto difícil implementar esta teoría, aunque existen buenos trabajos relacionados como los de Marcu (1997) donde se propone un analizador retórico que es usado en aplicaciones tales como el resumen automático por filtrado.

Otros trabajos relacionados con la búsqueda de relaciones textuales son los *Sistemas de Exploración Contextual*. Aquí se puede mencionar el *Método de Exploración Contextual (MEC)* desarrollado por el grupo La.L.I.C.¹, dirigido por Jean Pierre Desclés (1996) y (1997), que provee el marco necesario para identificar información semántica específica contenida en los textos. El método se propone encontrar unidades textuales a las que denomina *indicadores* e *índices*. Los indicadores son expresiones lingüísticas que “disparan” la ejecución de un cierto número de reglas de exploración encargadas de determinar el sentido o el significado de una palabra o expresión. La aplicación de una de estas reglas examina el contexto del indicador buscando los índices lingüísticos complementarios de este.

3. REPRESENTACIÓN

Partiendo de la localización de marcadores y del estudio de su alcance, en esta sección mostramos una estructura que expresa la relación encontrada vinculando oraciones y párrafos. Tal estructura puede verse utilizando *XML (eXtended Markup Language)*, ya que este metalenguaje ofrece un excelente formato para describir datos estructurados. Como tiene la particularidad de que no posee etiquetas prefijadas, el diseñador las genera a su antojo de acuerdo con la necesidad de lo que se intenta describir.

La idea es construir un árbol asociado al texto (Figura 1) y conformar así una estructura que soporte los textos escritos en función de la localización de un conjunto de términos y el ámbito de incidencia de estos a lo largo del texto.

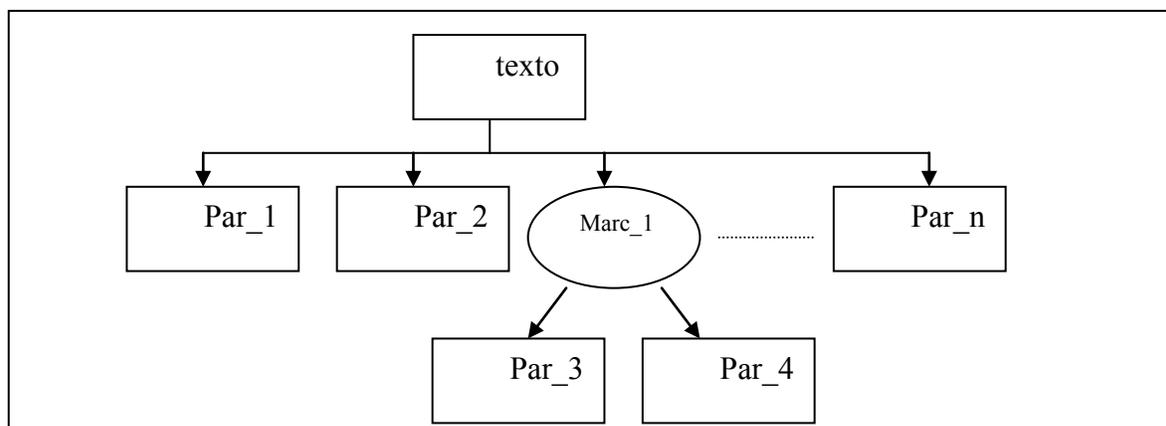


Figura 1. Grafo de la estructura del texto

¹ *Langage, Logique, Informatique, Cognition et Communication*, París IV, Francia.

Este objetivo supone la necesidad de contar con nodos de distinto tipo y características, puesto que, por un lado, existirán porciones de texto que no estén vinculadas al alcance de ningún marcador y que será necesario reflejar de manera que aparezca representada la totalidad del texto. Y, por otro lado, existirán porciones que sí estarán afectadas directamente por tales elementos. Si bien se pretende contemplar una estructura global que abarque todo el texto analizado, queda claro que en un texto donde no aparezcan marcadores discursivos, no tendrá sentido realizar un análisis de este tipo. En la Figura 1, el nodo rotulado *Marc_1* es binario y su alcance lo constituyen los párrafos 3 y 4 (nodos *Párr_3* y *Párr_4*).

Se propone una estructura que completa la información de los nodos correspondientes a marcadores del discurso con lo siguiente:

Nombre: contiene el nombre del marcador.

Etiqueta: etiqueta con que se identifica el marcador.

Ubicación: {al comienzo de párrafo (P), al comienzo de oración (O), intraoracional (I)}.

Tipo: {unario (U), binario (B), compuesto(C)}.

Clasificación: conector aditivo, estructurador ordenador, etc.

Nro_párrafo: contiene el número de párrafo al que pertenece el marcador (desde el comienzo del texto).

Nro_oración: contiene el número de oración a la que pertenece el marcador (desde el comienzo del texto).

Nro_desde_1: contiene el número de oración o párrafo (desde el comienzo del texto) donde comienza el alcance del 1er. miembro del marcador.

Nro_hasta_1: contiene el número de oración o párrafo (desde el comienzo del texto) donde termina el alcance del 1er. miembro del marcador.

Nro_desde_2: contiene el número de oración o párrafo (desde el comienzo del texto) donde comienza el alcance del 2do. miembro del marcador.

Nro_hasta_2: contiene el número de oración o párrafo (desde el comienzo del texto) donde termina el alcance del 2do. miembro del marcador.

Cada marcador definirá un subárbol con raíz en él y cuyos hijos estén conformados por el texto comprendido en el alcance de dicho marcador delimitado con etiquetas, que son las que especifican la relación estructural que se puede encontrar entre porciones de un texto.

3.1 Las reglas contextuales

Para el etiquetado aplicamos dos tipos de reglas: (1) aquellas que determinan los marcadores y asignan las etiquetas a los mismos (*operadores*); y (2) aquellas que etiquetan el texto –los párrafos, las oraciones y grupos de ellos– en función de los marcadores y su ubicación; es decir, marcan los *argumentos* establecidos por el alcance de dichos marcadores.

Para representar las reglas anteriores, se utiliza un formalismo (*Context Rules*) desarrollado por el Grupo de Procesamiento del Lenguaje Natural (PLN) del In.Co.². Su propósito es el análisis y etiquetado de porciones de texto que cumplen una determinada función dentro de este, considerando además que tales segmentos se encuentren en un cierto contexto. Las reglas incluyen palabras, signos de puntuación y las propias etiquetas asociadas a porciones de texto ya marcadas. La secuencia de elementos no tiene por qué ser contigua; pueden intercalarse “zonas” de las cuales solo se especifica un largo máximo y un conjunto de elementos no permitidos. Además, cualquier elemento que aparece en un texto puede considerarse para determinar si cierta información semántica está presente.

Una regla contextual es una expresión que identifica una porción de texto en un documento y le coloca una etiqueta. En esta regla, se puede expresar cierto contexto a la izquierda y/o a la derecha de la porción analizada que determine efectivamente la función que cumple el segmento analizado. Se establece, entonces, un cierto “orden” en el cual se deben presentar las distintas unidades en el texto. Dichas unidades pueden no aparecer contiguas y distar un máximo (medido en *tokens*) entre una y otra, así como indicar que determinados elementos no han de estar presentes entre ellas (zona de exclusión).

Sintaxis básica:

Etiqueta → *Contexto izquierdo* \ *Cuerpo* / *Contexto derecho* ; *Especificación de conjunto*,
donde:

- *Etiqueta*: es el *tag* asociado a la porción de texto contenida en *Cuerpo*.
- *Contexto izquierdo*, *Cuerpo*, *Contexto derecho*: secuencia de elementos pertenecientes a la denominada *zona de exclusión* y *etiquetas*.
- \ , /: marcas para delimitar cual es la porción analizada y sus contextos (izquierdo y derecho). Los contextos pueden ser vacíos (incluso ambos), pero el cuerpo no.
- La *zona de exclusión* es una expresión de la forma **(Conj_excl, máximo)*, donde *Conj_excl* es el nombre de un conjunto de etiquetas y *máximo* es un número natural.

² Instituto de Computación de la Facultad de Ingeniería, Uruguay.

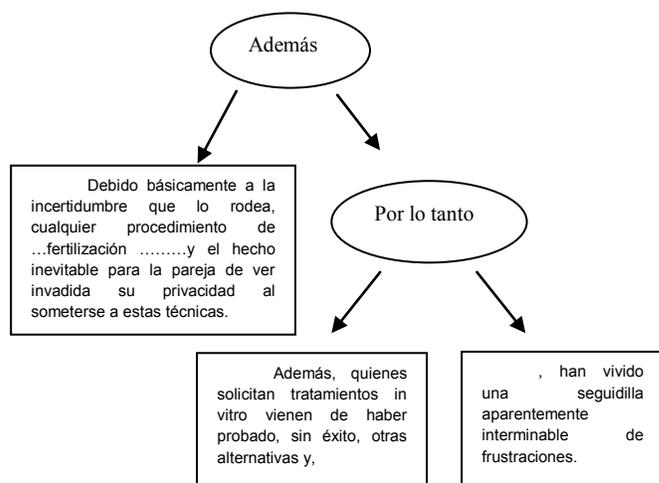
- *Especificación de conjunto*: es la definición por enumeración de los elementos especificados en la zona de exclusión. Puede ser vacío.

El formalismo completo de estas reglas puede encontrarse en Wonsever y Minel (2001). Por otra parte, en este trabajo pretendemos, además, especificar la estructura del texto y modelar este formalismo con XML, aspecto que abordaremos en la siguiente sección, particularmente para el marcador *además*.

4. CASO DE ESTUDIO: EL CONECTOR ADEMÁS

En esta sección, mostramos el formalismo aplicado al conector *además* en discursos escritos. El marcador puede encontrarse tanto al comienzo de una oración como dentro de esta (*intraoracional*). Este último caso no será considerado en este trabajo³.

Cuando el marcador es la primera palabra de la oración y, a su vez, esta es la primera oración de un párrafo, el alcance es a nivel de párrafos; es decir, el segundo miembro será el párrafo que contiene al marcador, mientras que el primer miembro lo constituirá el párrafo inmediato anterior. Si fuera la primer palabra de una oración y esta no fuera la primera del párrafo, los miembros que constituyen el alcance del marcador son la oración que lo contiene (2do. miembro) y la anterior a esta (1er. miembro)⁴.



(3) *Debido básicamente a la incertidumbre que lo rodea, cualquier procedimiento de fertilización asistida desencadena, de manera casi inevitable, distintos niveles de estrés ansiedad y angustia. Hay muchas fuentes...*

*Además, quienes solicitan tratamientos in vitro vienen de haber probado, sin éxito, otras alternativas y, **por lo tanto**, han vivido una seguidilla aparentemente interminable de frustraciones.*

Figura 2. Grafo con marcador al comienzo de párrafo

³ Más allá de que existen ocasiones de que si bien el término es intraoracional, el miembro que contiene el marcador es presentado como un agregado a lo expuesto en un contexto anterior manteniendo una escala argumentativa. En estos casos estamos en presencia de incisos.

⁴ Esto no siempre es así en todos los marcadores estudiados; dependerá de cada caso.

En el Ejemplo (3) aparecen los marcadores: además y por lo tanto. Si bien no se hace un análisis más detallado en este trabajo de las ocurrencias intraoracionales, ya que por lo general su alcance abarca la oración que contiene al marcador –y acá se pretende mostrar estructuras discursivas que van más allá de una oración-, se muestra parte del grafo asociado en la Figura 2 que expresa como debiera ser su representación y la Tabla 1 muestra la estructura con la información asociada.

NOMBRE	Además
ETIQUETA	<ADEMAS_P>
UBICACIÓN	P
TIPO	B
CLASIFICACION	Conector aditivo
NRO_PARRAFO	2
NRO_ORACION	3
NRO_DESDE_1	1
NRO_HASTA_1	1
NRO_DESDE_2	2
NRO_HASTA_2	2

Tabla 1. Estructura completa

Una característica del término *además* es que este puede cumplir tanto la función de conector aditivo como así también la de reforzar argumentativamente una porción de texto como refleja el Ejemplo (4):

- (4) La demanda fue presentada por Corrado Ferlaino, delegado administrador y presidente de la SA Napoli, poseedora del 93,5% de las acciones del club, que reclama **además** una indemnización de unos US\$ 125 millones...

En estos casos, el *además*, que no aparece entre comas no se debe detectar como marcador discursivo ya que su función es la de reforzar el argumento (en este caso, el reclamo. Es necesario manejar hipótesis para desambiguar este tipo de ocurrencias. Un ejemplo es el que recogemos en la Tabla 2. Habremos de contar con reglas que especifiquen este tipo de restricciones.

Hipótesis: Marcador entre signos de puntuación

Para que cumpla la función discursiva, debe aparecer:

- entre comas si es intraoracional;
- si ocupa el primer lugar en una oración, debe venir seguido de la coma.

Tabla 2. Hipótesis de desambiguación

4.1. Las reglas para además

A renglón seguido, representamos la regla para el caso del marcador *además* al comienzo de párrafo.

```
arg1_ademas_p → <P><O> \ *(S,500) / </O></P> <P><O><ADEMAS> CM ;
S={</P>}5
```

```
arg2_ademas_p → <P><O><ADEMAS> CM \ *(S,500) / </O></P>
```

```
gr_adit_ademas_p → \<P><O> arg1_ademas_p </O></P><P><O><ADEMAS>CM
arg2_ademas_p </O></P> /
```

En las reglas precedentes, aparece la siguiente simbología:

<P>: párrafo.

<O>: oración.

<CM>: puntuación (coma o punto y coma).

arg1_ademas_p: primer argumento del conector de párrafo además.

arg2_ademas_p: segundo argumento del conector de párrafo además.

gr_adit_ademas_p: estructura compleja con ambos argumentos.

Como forma de obtener una DTD (Document Type Definition) para dichas reglas se podría tener:

```
<!DOCTYPE gr_adit_ademas_p [
<!ELEMENT gr_adit_ademas_p (arg1_ademas_p,arg2_ademas_p)>
<!ELEMENT arg1_ademas_p (P,O,arg1)>
<!ELEMENT arg1 (#PCDATA)>
<!ELEMENT arg2_ademas_p (P,O,Ademas_p,arg2)>
<!ELEMENT arg2 (#PCDATA)>
<!ELEMENT Ademas_p (#PCDATA)>
```

⁵ Con la expresión $S=\{\langle/P\rangle\}$, se indica que en la zona de exclusión, no puede aparecer un marcador de cierre de párrafo. Es decir que el alcance del primer miembro es justamente hasta que aparezca el marcador además como primer palabra del nuevo párrafo.

```
<!ATTLIST Ademas_p marcador CDATA #REQUIRED, ubicacion (P|O|I)“O”,
clasificacion CDATA #REQUIRED, nro_parrafo CDATA #REQUIRED, ... )> ]>
```

Utilizamos un analizador basado en el formalismo explicado en la sección 3.1. Una vez aplicado dicho analizador (basado en Reglas Contextuales), el Ejemplo (3) etiquetado con las reglas especificadas queda de la siguiente manera:

```
<gr_adit_ademas_p> <arg1_ademas_p><P Num="1"><O Num="1">
<arg1> Debido básicamente a la incertidumbre que lo rodea, cualquier
procedimiento de fertilización asistida desencadena, de manera casi
inevitable, distintos niveles de estrés, ansiedad y angustia.</O><O Num="2">
Hay muchas fuentes de agobio:el impacto, ... </O> </arg1> </P>
</arg1_ademas_p><arg2_ademas_p> <P Num="2"> <O
Num="3"><Ademas_p marcador="Además," ubicacion="P",
clasificacion="conector aditivo"...></Ademas_p><arg2>Además, quienes
solicitan tratamientos in vitro vienen de haber probado, sin éxito, otras
alternativas y <Por_tanto_i>, por lo tanto,</Por_tanto_i> han vivido una
seguidilla aparentemente interminable de frustraciones. </O> </arg2> </P>
</arg2_ademas_p> </gr_adit_ademas_p>
```

4.2. Interacción con otros marcadores

Cuando se amplía el conjunto de marcadores estudiados –es decir, se agregan elementos de distinto tipo (según la clasificación en la que se basa este estudio)–, es necesario definir otro conjunto de reglas para determinar “prioridades” en cuanto al alcance de cada marcador, debido a la posible superposición de sus miembros. Establecemos, pues, una jerarquía entre los distintos marcadores. Por ejemplo, ordenadores por encima de los conectores aditivos; estos, a su vez, por sobre los contraargumentativos; estos por sobre los reformuladores, entre otras.

5. RESULTADOS OBTENIDOS

Para la implementación, usamos un analizador sintáctico para reglas contextuales⁶. Trabajamos con un conjunto de párrafos de la Real Academia Española (Corpus *CREA*⁷), así como también párrafos de noticias de los diarios de Montevideo, Uruguay. Consideramos más de cuatro mil quinientos párrafos para derivar las reglas. Para la

⁶ Desarrollado en el Grupo PLN-InCo.

⁷ *Corpus de Referencia del Español Actual*.

validación de los criterios de alcance, contamos con setenta documentos completos (y diferentes a los empleados para la formulación de las reglas) a los efectos de comprobar mejor la incidencia que pudieran tener a lo largo del documento, así como la interacción de unos marcadores sobre otros.

Si utilizamos las medidas estándar en recuperación de información, *precision* y *recall*, tomando en cuenta el marcador *además*, nos percatamos de que los resultados obtenidos son muy buenos si bien se trata de una pequeña muestra.

$$\textit{Precision} \text{ (ruido)} = \text{recuperados correctos} / \text{recuperados} = 25 / 25 = 1$$

$$\textit{Recall} \text{ (silencio)} = \text{recuperados correctos} / \text{total de correctos} = 25 / 28 = 0.89$$

Cuando estas medidas se llevan a más de los marcadores trabajados:

$$\textit{Precision} = \text{recuperados correctos} / \text{recuperados} = 246 / 262 = 0.94$$

$$\textit{Recall} = \text{recuperados correctos} / \text{total de correctos} = 246 / 286 = 0.86$$

6. CONCLUSIONES

El presente artículo se enmarca en el análisis del discurso escrito y tiene por objeto la búsqueda de ciertas unidades lingüísticas que determinan la presencia de relaciones entre segmentos textuales. Para ello, atendemos a su ubicación en el texto, cuál es la porción afectada por cada marcador, y proponemos una estructura de todo el texto en función de las relaciones cuyo centro es el marcador.

Al realizar la validación de resultados, comprobamos que los marcadores eran generalmente reconocidos. Cabe destacar que este análisis cobra sentido para aquellos documentos que poseen un buen número de marcadores discursivos.

Como línea de investigación futura, pretendemos aumentar el número de marcadores estudiados en este momento estamos manejando en el orden de los treinta términos—, a la vez de avanzar en el análisis de las precedencias entre los distintos términos (superposición de miembros).

Una aplicación concreta donde se puede aprovechar este tipo de análisis es el resumen automático por filtrado. Existe ya algún trabajo sobre este tema para el español (González, 2005). En estos casos, es importante mantener no solo la idea temática de lo que se está resumiendo, sino también la mayor coherencia posible en las porciones de texto que son extraídas. Así mismo, es conveniente tener determinada la ubicación de la palabra en una oración o párrafo, de manera tal que, al extraer un segmento que lo contenga, sea necesario obtener también el contexto determinado por su alcance.

Referencias

- CAVIGLIA, Serrana, MALCUORI, Marisa y GRASSI, Mariela, (2000), «Corpus Informatizado: textos del español del Uruguay (CORIN)», *IV Congreso de Lingüística General*, Universidad de Cádiz, abril de 2000.
- DESCLÉS, Jean Pierre (1996), «Systèmes d'exploration contextuelle», *Actes du colloque sur le Calcul du sens et contexte*, Caen: Université de Caen.
- DESCLÉS, Jean Pierre, CARTIER, Emmanuel, JACKIEWICZ, Agata y MINEL, Jean Luc (1997), «Textual Processing and Contextual Exploration Method», *Context 97*, Río de Janeiro. Disponible en: <http://lalic.paris-sorbonne.fr/PUBLICATIONS/1997-1998/Descles/Rio97.pdf> [Consultado el 23 de febrero de 2012].
- EXTENSIBLE MARKUP LANGUAGE (XML) 1.0 (Second Edition, 2000). Disponible en <http://www.w3.org/TR/REC-xml> [Consultado el 23 de febrero de 2012].
- FIGUERAS, Carolina (1999), «Puntuación y conectores causales», *Espéculo. Revista de estudios literarios* 13. Disponible en http://www.ucm.es/info/especulo/numero13/punt_cc.html [Consultado el 23 de febrero de 2012].
- GONZÁLEZ, José E. y CLAVELL, Roberto (2005), *Resumidor de textos basado en marcadores discursivos*, proyecto de grado. Disponible en <http://www.fing.edu.uy/inco/grupos/pln/prgrado/InformePGResumen.pdf> [Consultado el 23 de febrero de 2012].
- MANN, William C. y THOMPSON, Sandra A. (1988), «Rhetorical Structure Theory: Toward a functional theory of the text organization», *Text* 8(3): 243-281. Disponible en <http://www.cis.upenn.edu/~nenkova/Courses/cis700-2/rst.pdf> [Consultado el 23 de febrero de 2012].
- MARCU, Daniel (1997), *The Rhetorical Parsing, Summarization and Generation of Natural Languages Texts*, tesis doctoral, Toronto: Universidad de Toronto.
- MARTÍN ZORRAQUINO, M. A. y PORTOLÉS, J. (1999), «Los marcadores del discurso», en I. BOSQUE y V. DEMONTE (eds.), *Gramática Descriptiva de la Lengua Española*, tomo III, Madrid: Espasa Calpe, pp. 4051-4203.
- MINEL, Jean Luc, NUGIER, Sylvaine y PAIT, Gérald (1997), «How to appreciate the quality of automatic text summarization? Examples of FAN and MLUCE protocols and their results on SERAPHIN», *EACL97, Intelligent Scalable Text Summarization*, pp. 25-30. Disponible en <http://www.aclweb.org/anthology-new/W/W97/W97-0705.pdf> [Consultado el 23 de febrero de 2012].
- PRADA, Juan José (2001), *Marcadores del discurso en español. Análisis y representación*, tesis de maestría. Disponible en <http://www.fing.edu.uy/inco/pedeciba/bibliote/tesis/tesis-prada.pdf> [Consultado el 23 de febrero de 2012].
- REAL ACADEMIA ESPAÑOLA (1973), *Esbozo de una nueva gramática de la lengua española*, Madrid: Espasa Calpe.

- ROSSARIE, Corinne (2000), *Connecteurs et relations de discours: des liens entre cognition et signification*, Nancy: Presses Universitaires de Nancy II.
- WONSEVER, Dina y MINEL, Jean Luc (2001), «Contextual Rules for Text Analysis», *Lecture Notes in Intelligence Artificial*, pp. 509-523. Disponible en <http://halshs.archives-ouvertes.fr/docs/00/09/77/97/PDF/Cicling.pdf> [Consultado el 23 de febrero de 2012].